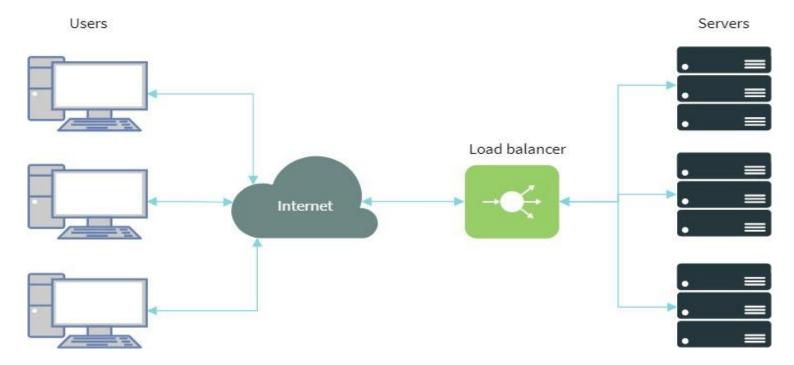
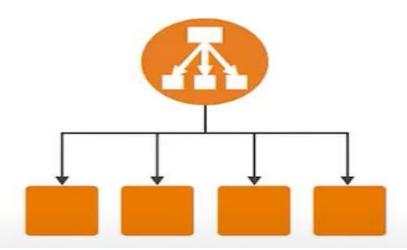
An Elastic Load Balancer is a service which uniformly distributes network traffic and workloads across multiple servers



Types of Load Balancer

- Classic Load Balancer
- Application Load Balancer
- Network Load Balancer
- Gateway Load Balancer

 A Classic load balancer distributes equally incoming application traffic across multiple EC2 instances in multiple Availability Zones. This increases the fault tolerance of your applications. Elastic Load Balancing detects unhealthy instances and routes traffic only to healthy instances.



- Create Linux EC2 Machine
- Bootstrap Script Code: Click Here
- Enable HTTP Port
- Create Second Linux EC2 Machine
- Bootstrap Script Code: Click Here
- Enable HTTP Port

Create Load balancer

Load Balancing



Load Balancer

- Click on Create for Classic Load Balancer
- Define Load Balancer
- Assign Security Group
- Configure Security Settings

- Configure Health Check (Important)
- Enter Response timeout as 2
- Enter Interval as 5
- Enter Unhealthy threshold as 2
- Enter healthy threshold as 2

- Add EC2 Instances
- Add Tags
- Review
- Click on Create
- Copy DNS Name & Paste in the browser
- Stop First EC2 Machine

Quotas for your Classic Load Balancer:

Document: <u>Click Here</u>

Your AWS account has the following quotas related to Classic Load Balancers.

- Load balancers per Region: 20
- Listeners per load balancer: 100 †
- Security groups per load balancer: 5
- Registered instances per load balancer: 1,000
- Subnets per Availability Zone per load balancer: 1 †

† These quotas cannot be increased.

- The load balancer also monitors the health of its registered targets and ensures that it routes traffic only to healthy targets. When the load balancer detects an unhealthy target, it stops routing traffic to that target. It then resumes routing traffic to that target when it detects that the target is healthy again.
- Example. You enter a retail store, pick up things you need to buy and approach the billing counters. You see, there are 3 counters that are open and all 3 of them has a very long queue. The store manager sees you and other customers getting annoyed. He decides to open the other two counters. Now the load on those three counters gets reduced and eventually divided amongst the five counters. This makes the customers happy and reduces the cashier's workload. This is the concept of load balancing.

- **Response Timeout**: The amount of time to wait when receiving a response from the health check, in seconds. Valid values: 2 to 60
- Interval: The amount of time between health checks of an individual instance, in seconds. Valid values: 5 to 300
- Unhealthy Threshold: The number of consecutive failed health checks that must occur before declaring an EC2 instance unhealthy. Valid values: 2 to 10
- Healthy Threshold: The number of consecutive successful health checks that must occur before declaring an EC2 instance healthy. Valid values: 2 to 10